Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth

Ralf Steuer1,*, Henning Knoop1 and Rainer Machné2

1 Institute of Theoretical Biology, Humboldt-University Berlin, Invalidenstr. 43, D-10115 Berlin, Germany
2 Institute for Theoretical Chemistry, University of Vienna, Austria

* To whom correspondence should be addressed. E-mail: ralf.steuer@hu-berlin.de

Received 18 October 2011; Revised 12 January 2012; Accepted 13 January 2012

Abstract

Cyanobacteria are phototrophic microorganisms of global importance and have recently attracted increasing attention due to their capability to convert sunlight and atmospheric CO2 directly into organic compounds, including carbon-based biofuels. The utilization of cyanobacteria as a biological chassis to generate third-generation biofuels would greatly benefit from an increased understanding of cyanobacterial metabolism and its interplay with other cellular processes. In this respect, metabolic modelling has been proposed as a way to overcome the traditional trial and error methodology that is often employed to introduce novel pathways. In particular, flux balance analysis and related methods have proved to be powerful tools to investigate the organization of large-scale metabolic networks—with the prospect of predicting modifications that are likely to increase the yield of a desired product and thereby to streamline the experimental progress and avoid futile avenues. This contribution seeks to describe the utilization of metabolic modelling as a research tool to understand the metabolism and phototrophic growth of cyanobacteria. The focus of the contribution is on a mathematical description of the metabolic network of *Synechocystis* sp. PCC 6803 and its analysis using constraint-based methods. A particular challenge is to integrate the description of the metabolic network with other cellular processes, such as the circadian clock, the photosynthetic light reactions, carbon concentration mechanism, and transcriptional regulation—aiming at a predictive model of a cyanobacterium *in silico*.

Key words: Ecosystems biology, flux balance analysis (FBA), network reconstruction, photosynthesis, quantitative modelling, systems biology.

Introduction

Cyanobacteria are phototrophic microorganisms and the only known prokaryotes capable of oxygenic photosynthesis. Having evolved possibly as early as 3.5 billion years ago (Des Marais, 2000), cyanobacteria are of global importance for almost all geochemical cycles and had profound impact on life on Earth. Today, cyanobacteria are still responsible for a significant fraction of global primary productivity and remain major players in global oxygen supply, carbon sequestration, and nitrogen fixation.

From a metabolic perspective, cyanobacteria are highly versatile organisms and occupy almost every environment where light is available. Renewed attention on the organization of cyanobacterial metabolism is driven by their capability to convert sunlight and atmospheric CO2 directly into carbon-based compounds, therefore providing an opportune biological chassis to generate third-generation biofuels (Atsumi *et al.*, 2009; Ducat *et al.*, 2011; Hess, 2011; Quintana *et al.*, 2011). Indeed, biofuels derived from cyanobacteria offer several potential advantages, as compared with fuels derived from land-based plants or other renewable sources. Cultivation of cyanobacteria does not require large amounts of arable land and is therefore not in direct competition with agricultural food production. Rather, cyanobacteria can be grown in large quantities relying only on water, including salt water, some minerals, CO2, and sunlight. Several cyanobacteria also possess the capability to fix atmospheric nitrogen, with the potential to reduce costs of fertilizer that may otherwise significantly affect the economic...
Computational models of metabolism

Computational modelling is increasingly recognized as an expedient research tool to understand the organization of biological systems. Nonetheless, the methods and practices of mathematical modelling are highly diverse and no single methodology alone is able to cover the diverse temporal and spatial scales observed in biological systems. Therefore, the future of modelling resides in the utilization of a combination of methods, each suited to describe a particular aspect of biological reality—giving rise to the challenge to combine these diverse conceptual and computational pictures into a coherent whole.

Common to almost all methods of modelling is that they seek to translate a given biological process into a formal language. According to the view put forward here, this translation of biological reality into the mathematical realm serves two distinct purposes (Steuer and Junker, 2009; Steuer, 2011): first, the translation of an assumed functional interaction into a mathematical representation allows researchers to communicate knowledge in a way that is not possible by mere verbal description. Paradigmatic examples are already provided by simple enzyme kinetics for which the formulation of models allows communication of complex patterns of interactions with only a few parameters. By describing assumed molecular mechanisms, such as competitive inhibition, using a mathematical representation allows other researchers to contrast their own results in a precise and well-defined way. Secondly, once the interactions underlying a biological process are cast into a mathematical form, mathematical theory and computational methods provide powerful tools to predict the emergent outcome of any particular set of interactions—again significantly surpassing the scope of human intuitive reasoning. A paradigmatic example is the emergence of oscillatory behaviour in biochemical reaction networks. Once translated into an appropriate mathematical representation, the answer to the question of whether a given reaction mechanism allows for sustained oscillations is little more than a technical formality. Mathematical modelling therefore allows the study of the consequences and outcomes of assumed interactions using computational or analytical techniques. As will be emphasized below, both of these aspects stand on their own and are equally important to understand the functioning of biological systems. Importantly, the use of a mathematical language, in the sense outlined above, does not pre-suppose or require trueness of the mathematical description.

While most techniques of modelling apply in a similar way to a variety of cellular processes, the focus of this review is foremost a description of models of cyanobacterial metabolism. In many aspects, metabolic networks of prokaryotic organisms are specifically suited for computational modelling. This advantage can be attributed to three characteristics that distinguish most prokaryotic metabolic networks from other networks of cellular interactions (Steuer, 2011): first, the function of many biochemical pathways resides in the mass transfer and synthesis of metabolic compounds. This fact can be exploited to resolve the topological organization of metabolic networks, for example by isotopic labelling techniques. In contrast, a similar technique that is able to trace functioning pathways does not exist for cellular information processing networks. Likewise, biochemical reaction networks are subject to constraints with respect to physicochemical and thermodynamic feasibility. Correspondingly, the topology of central metabolism is reasonably well understood, at least for many prokaryotic organisms. Secondly, the modes of action of the building blocks of metabolic networks, enzymes that catalyse biochemical reactions, have been studied for more than a century. Most enzymes are reasonably well described by conventional chemical kinetics and their action can often be replicated \textit{in vitro}. Therefore, the local dynamics of enzymatic reactions are reasonably well understood. Thirdly, and probably most important, the metabolic network of prokaryotic organisms can usually be assigned a well-defined functionality, namely the synthesis of cellular building blocks to support cellular growth and the provision of ATP and other ‘currency metabolites’ for cellular maintenance (Ibarra \textit{et al.}, 2002; Lewis \textit{et al.}, 2010). These rather clear-cut objectives of metabolic pathways form the basis of the success of the various optimization techniques applied to large-scale metabolic networks.
Although there are several caveats with respect to the arguments outlined above, a known topology, reasonably well understood modes of local action, and a clear-cut functionality offer a good starting point for the construction of models of cellular metabolism.

A hierarchy of computational description

In practice, metabolic models take a variety of forms and no single computational methodology is able to describe all possible aspects of metabolic functioning (Steuer and Junker, 2009). Rather, a hierarchy of mathematical description or frameworks exists, each with particular strengths and weaknesses. An overview is given in Fig. 1.

Currently, there is a dichotomy between small-scale models, usually involving a high level of detail with respect to the interacting compounds on the one hand and large-scale models of cellular networks on the other hand. Located on the right side within the continuum shown in Fig. 1 are detailed kinetic models of cellular pathways. Detailed kinetic models are usually implemented as a set of ordinary differential equations for all biochemical compounds within a pathway. Typically, the construction of such models makes use of aggregated biochemical rate laws, such as Michaelis–Menten kinetics, to describe biochemical interconversions of compounds. Multiple repositories and simulation tools exist that allow on- and offline analysis of kinetic models (Olivier and Snoep, 2004; Hoops et al., 2006; Machné et al., 2006; van Gend et al., 2007; Li et al., 2010). However, the availability of detailed kinetic models of cellular pathways, in particular for cyanobacteria or other phototrophic organisms, is still extremely limited. One reason for this scarcity is due to the fact that detailed kinetic models put extensive demands on data availability. Their construction usually requires explicit knowledge of all involved rate constants and kinetic parameters. For cyanobacteria, detailed kinetic models are mostly available for selected subsystems, such as the cyanobacterial circadian clock (Miyoshi et al., 2007; Rust et al., 2007; Cerveny and Nedbal, 2009; Brettschneider et al., 2010), photosystem II (Jablonsky and Lazar, 2008), carbon-concentrating mechanisms (Badger et al., 1985; Fridlyand et al., 1996), as well as for selected metabolic pathways, such as the Calvin-Benson cycle (Jablonsky et al., 2011). Kinetic models often also rely on kinetic parameters and data from plants and other related organisms. Small quasi-autonomous subsystems, such as the circadian clock, are particularly suited for kinetic modelling as they typically involve only few interacting partners that operate in a manner sufficiently isolated from other cellular networks and therefore allow a data-driven construction of detailed kinetic models.

Given the difficulties associated with the construction of detailed kinetic models, most current efforts towards modelling cellular metabolism are located on the left side within the continuum shown in Fig. 1. Topological methods are understood here as methods that represent cellular metabolism as a, usually bipartite, graph. Such graph-based descriptions of metabolism, while popular in many areas of interdisciplinary sciences (Barabási and Oltvai, 2004), usually do not incorporate genuine biochemical quantities, such as reaction stoichiometries, in the description of the network—a shortcoming that significantly limits their potential to resolve genuine biological questions (Steuer et al., 2005).

Fig. 1. A hierarchy of computational representations. Current approaches to metabolic modelling range from graph-based methods to detailed kinetic models of cellular pathways. Methods that aim at large-scale description are typically less quantitative. Of particular interest are also intermediate methods that seek to bridge the gap from detailed kinetic to large-scale models of cellular metabolism. Adapted and redrawn from Steuer R and Junker BH. 2009. Computational models of metabolism: stability and regulation in metabolic networks. In: Advances in chemical physics. This material is reproduced with permission of John Wiley & Sons, Inc.
A better choice for the description of biochemical networks are usually stoichiometric models. Stoichiometric models explicitly take the physico-chemical nature of biochemical reaction networks into account, but their analysis does not yet require kinetic parameters or detailed knowledge of the properties of the involved enzymes. Prominent applications of stoichiometric modelling are the concept of elementary flux modes (EFMs), as well as FBA. In particular, FBA has become an important tool to analyse the properties of large-scale metabolic networks. FBA is entirely based on constraints imposed by mass conservation and is able to incorporate thermodynamic (Beard et al., 2004; Kümmel et al., 2006; Henry et al., 2007; Hoppe et al., 2007) and other constraints (Zhuang et al., 2011).

Of particular interest are also intermediate methods that seek to bridge the gap from stoichiometric to explicit kinetic models of cellular pathways. Intermediate methods may, for example, take the form of hybrid models, coupling a small-scale kinetic model to a large-scale stoichiometric model of metabolism (Mahadevan et al., 2002; Luo et al., 2006), or aim to derive dynamic properties of large reaction networks without requiring detailed knowledge of kinetic parameters. The latter approach typically makes use of Monte Carlo sampling of the parameter space to obtain a probabilistic understanding of the dynamic properties of the respective system (Wang et al., 2004; Steuer et al., 2006; Tran et al., 2008; Steuer, 2011). It is expected that such methods will be of importance for the further development of models of phototrophic metabolism. In particular, several characteristic properties of cyanobacterial metabolism differ from their counterparts in conventional heterotrophic bacteria. These differences include the utilization of light as a resource, which necessitates the consideration of the fast and intrinsically dynamic light reactions as part of the metabolic network. Likewise, several dominant metabolic processes, such as the costs of uptake and cycling of inorganic carbon (Tchernov et al., 2001, 2003), as well as the inhibition of Rubisco by molecular oxygen, are highly dependent on concentrations and their gradients—and therefore difficult to capture within a framework predominantly based on stoichiometric relationships.

**Fig. 2.** Gene–protein–reaction associations for a selected subset of genes of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. Enzymes may be composed of several subunits and single genes may encode multifunctional enzymes.
**A stoichiometric reconstruction of *Synechocystis* sp. PCC 6803**

Large-scale stoichiometric modelling of metabolic networks relies entirely on a high-quality reconstruction of the underlying set of reactions. The aim of such a metabolic reconstruction is to give a comprehensive account of all biochemical conversions taking place within a living cell or organism, including transport and non-enzymatic reactions. The reconstruction process itself increasingly follows standard procedures and is described in detail in the literature (Feist *et al.*, 2009; Thiele and Palsson, 2010a). Here, the focus is on a recently published reconstruction of the cyanobacterium *Synechocystis* sp. PCC 6803 (Knoop *et al.*, 2010). The reconstruction process typically comprises four steps. The starting point is the annotated genome, as obtained from the CyanoBase Web site (Nakamura *et al.*, 1998, 1999; Nakao *et al.*, 2010), together with pathway repositories such as the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.kegg.com/; Kanehisa *et al.*, 2006) or MetaCyc (www.metacyc.org; Caspi *et al.*, 2010). These resources are combined to provide an initial draft network that consists of gene–protein–reaction associations for each annotated gene, in addition to a number of non-enzymatic (spontaneous) reactions and transport processes. A small-scale example is shown in Fig. 2, and the corresponding stoichiometric matrix is shown in Fig. 3.

Based on the draft reconstruction, an iterative process of gap finding and network curation begins. In particular, the draft network allows verification of whether all known metabolic intermediates of the respective organism can be synthesized using the set of reactions assigned to the organism. Usually, this is not the case and additional reactions have to be included in the model. For example, for the reconstruction of *Synechocystis* sp. PCC 6803 (Knoop *et al.*, 2010), the draft network did not support the synthesis of the amino acids glycine, serine, cysteine, methionine, asparagine, and histidine. To complete the network and to account for missing steps, different strategies may be followed. To this end, an increasing number of algorithms is available that aim to identify missing steps within metabolic networks computationally and offer automated schemes to provide suggestions for missing enzymes (Kumar *et al.*, 2007). These algorithms are usually based on shortest-path or minimal extension criteria to complete the metabolic network and to establish a defined metabolic functionality. However, it needs to be emphasized that such automated schemes do not always lead to meaningful results, owing to the fact that nature itself often does not follow a logic of minimal extensions but may choose seemingly capricious and non-optimal solutions for metabolic interconversion routes. Therefore, in particular for prokaryotic networks of smaller size, manual curation remains the most important step in metabolic network reconstruction. The knowledge accumulated in the biochemical literature on selected pathways often provides a crucial resource to re-annotate enzymes and to provide a correct picture of biochemical interconversions. For *Synechocystis* sp. PCC 6803, an example of the necessity of incorporating literature knowledge is given by the detailed

---

**Fig. 3.** The stoichiometric matrix corresponding to the gene–reaction associations given in Fig. 2. Columns correspond to reactions, and rows to metabolites.
elucidation of the photorespiratory pathways (Eisenhut et al., 2008) that were, at the time of reconstruction, not part of any pathway database. It is noted that tests for completeness and subsequent gap filling also must take into account various non-enzymatic reactions, including possible degradation reactions.

From network to model

Once the reconstruction can account for all relevant synthesis routes of metabolic intermediates, the network is converted into a mathematical model. This step involves the inclusion of several pseudo-reactions that account for cellular maintenance and growth, as well as the conversion of the set of reactions into a computer-readable format. Cellular maintenance is usually accounted for by adding an additional ATP drain. However, utilization of NADPH, independent of the synthesis reactions, and dilution of metabolites by growth may also be considered. Of particular importance is the biomass objective function (BOF) that represents an overall pseudo-reaction to describe cellular growth in terms of the required metabolic precursors. It is assumed that for growth and replication a specified set of metabolic precursors needs to be synthesized in fixed stoichiometric ratios. The BOF therefore describes a cellular demand that the remaining synthesis reactions have to satisfy in order to achieve cellular growth. It is noted that the BOF may change depending on experimental conditions. For cyanobacteria, different BOFs are usually assigned to heterotrophic, mixotrophic and phototrophic conditions, reflecting differences in cellular composition (Shastri and Morgan, 2005). It is noted that a mathematical model differs from a mere list of reactions in its scope to assign a particular functionality to any set of reactions, such as the capability to synthesize certain metabolic intermediates.

Finally, given a mathematical representation of the network, preferably using a defined exchange format such as SBML (Systems Biology Markup Language; Hucka et al., 2003), the model can be investigated using standard methods for stoichiometric analysis. To this end, an increasing number of software packages are available to facilitate large-scale network analysis (Becker et al., 2007; Klamt et al., 2007; Hoppe et al., 2011; Schellenberger et al., 2011). The entire process is highly iterative, and is summarized in Fig. 4. It is noted that at this stage, the model itself is not expected to be a fully verified and error-free representation of the respective metabolic network. Rather, its purpose is to provide a starting point to allow systematic functional verification, to identify uncertain and possibly misannotated enzymatic steps, and therefore to guide further experimentation.

It should be emphasized that the standardized reconstruction process, followed by stoichiometric analysis, was mainly developed to describe heterotrophic growth of...
various microorganisms. For example, within a recent effort to generate automatically 130 genome-scale metabolic models representing a taxonomically diverse set of bacteria (Henry et al., 2010), only a single cyanobacterium was included. A conspicuous lack of phototrophic metabolic reconstructions was also noted in a recent overview on applications of genome-scale reconstructions (Öberhardt et al., 2009). Indeed, several aspects of phototrophic metabolism can, as yet, only be inadequately represented by stoichiometric models. For example, light is typically represented as a single metabolic compound (photon) that participates in stoichiometric reactions along with other metabolites. Only recently, a first attempt at a novel light modelling approach that allows resolution of wavelength and photon flux has been proposed (Chang et al., 2011).

Problems in network reconstruction

In practice, a considerable number of further difficulties complicate the process of network reconstruction outlined above. Even for well-annotated organisms, such as Synechocystis sp. PCC 6803, errors in gene annotation are not uncommon. Likewise, integration of multiple reaction repositories and literature research is considerably hampered by different and often inconsistent naming schemes of metabolic intermediates (Radrich et al., 2007). For example, within KEGG, the metabolite 2-oxoisovalerate (COMPOUND: C00141) is listed with no less than nine synonyms (3-methyl-2-oxobutanoic acid, 3-methyl-2-oxobutyric acid, 3-methyl-2-oxobutanoate, 2-oxo-3-methylbutanoate, 2-oxoisovalerate, 2-oxoisopentanoate, α-ketovaline, 2-ketovaline, and 2-keto-3-methylbutyric acid), each of which may also be used within the literature and may therefore confuse results when researching the corresponding bibliome. Adding to the complexity of metabolite annotation, metabolites usually also co-exist in different protonation states depending on the intracellular milieu—a fact that must be taken into account when reactions from different sources are combined. Such inconsistent annotation may lead to erroneous gaps within the network, as pathways may no longer be connected. While most of these issues are straightforwardly resolved using manual inspection and expert knowledge, the large size of many metabolic networks often requires automated solutions. To this end, researchers working on metabolic network reconstructions increasingly recommend the utilization of a controlled vocabulary that allows consistent naming of metabolites to facilitate cross-database comparison. Such a controlled vocabulary usually refers to databases and web repositories, such as ChEBI (Chemical Entities of Biological Interest), a freely available dictionary of metabolic compounds (de Matos et al., 2010). As a further advantage, consistent and computer-readable naming allows software tools to recognize the annotated compound and therefore allows automated quality checks on the reconstructed networks to be performed. Even if sourced from curated pathway databases, reactions may not always be balanced with respect to elements and charge, resulting in metabolic cycles that can generate metabolic compounds without input. For example, Yoshikawa et al. (2011) reported that they failed to simulate a previously published model of Synechocystis sp. PCC 6803 (Montagud et al., 2010) because the reconstruction contained inadequate reaction loops. It is therefore strongly recommended to perform automated charge and elemental balancing at the end of the reconstruction process, as, for example, facilitated by COBRA 2 (Schellenberger et al., 2011) or the freely available toolbox SuBliMinaL (Swainston et al., 2011).

It should be mentioned that during network reconstruction not all identified missing steps correspond to erroneous gaps within the respective metabolic network. Cyanobacteria are known for highly streamlined genomes, in particular the genus Prochlorococcus (Partensky and Garczarek et al., 2010). For example, it was recently shown that Prochlorococcus, whose genomes lack catalase and therefore should be highly susceptible to damage from hydrogen peroxide, are protected by an extant HOOH-consuming microbial community in the surface mixed layer of the oligotrophic ocean (Morris et al., 2011). Likewise, recently a globally distributed marine cyanobacterium, UCYN-A, was described that lacks the oxygen-producing photosystem II complex. Instead, UCYN-A exhibits a strongly restricted photofermentative metabolism and is probably dependent on other organisms for essential compounds (Tripp et al., 2010). In such cases, unsupervised automated reconstruction may yield erroneous solutions that account for metabolic pathways that are not present in the respective organism.

Network size and quality criteria

One of the most persistent problems in current network reconstructions is that clear-cut criteria whose properties constitute a good reconstruction are still missing. While stoichiometric consistency and other formal aspects can be tested rather straightforwardly, the reconstruction process itself still allows for miscellaneous choices of criteria for inclusion or exclusion of specific reactions. This shortcoming is particularly relevant for enzymes with unclear specificity. In this case, a single enzyme might be annotated for a large number of possible interconversions, whereas in practice the enzyme is often for a smaller, but unknown, subset of interconversions. Common cases also include generic annotations in which enzymes, for example, require a ‘reduced acceptor’ or ‘hydrogen donor’ (both KEGG Compound C00030). Since current flux balance software can usually only deal with specific metabolites, not classes of metabolites, such generic terms may lead to a combinatorial inflation of the number of possible reactions. A computational representation of enzymes acting on polydisperse substrates, such as many storage compounds, also remains a challenging task (Kartal et al., 2011). Likewise, cofactor requirements of most reactions are often unknown. For example, within the
reconstruction of *Synechocystis* sp. PCC 6803, many reactions are included as NAD/NADH- and NADP/NADPH-dependent variants, as the actual specificities of those reactions are unknown.

In general, a large number of current reconstructions opt for the inclusion of as many reactions as possible, even at the expense of overestimating the number of actual interconversions for most enzymes. On the other hand, several recent studies also deliberately only focus on a selected subset of reactions, usually central carbon metabolism and adjacent pathways. It was observed previously (Sweetlove and Ratcliffe, 2011) that both approaches often give rise to an almost identical number of active reactions; that is, the number of reactions that carry non-zero metabolic flux under any condition is usually similar. While both approaches have their merits, it must be emphasized that the size of a reconstruction is only rarely an indicator of quality and usually does not increase the predictive power of a model. Indeed large size, in particular for poorly characterized organisms, often indicates indiscriminate inclusion of vaguely annotated enzymes.

A way to overcome some of these problems again resides in consistent annotation. In this respect, a first step is to assign a level of confidence with each included reaction, such that other researchers can clearly identify on which evidence any particular reaction was included within the network. Such consistent annotation allows the discernment of whether a particular reaction was included on the basis of sound biochemical evidence, or whether, for example, it was included to restore a certain metabolic functionality whose absence would otherwise preclude further analysis. Likewise, there is increasing agreement that a complete reconstruction of any single organism and its validation is beyond the means of a single research group. Therefore, increasing emphasis is put on community efforts, such as network reconstruction jamborees that aim at a consensus reconstruction to reconcile and refine knowledge about the respective organism (Herrgård et al., 2008; Thiele and Palsson, 2010b). Such reconstruction jamborees were already successfully undertaken for several common organisms, including *Saccharomyces cerevisiae*, *Salmonella typhimurium*, and *Homo sapiens*, but, as yet, not for higher plants or cyanobacteria. Finally, successful reconstructions should be made available on dedicated exchange platforms, such as biomodels.org (Li et al., 2010) or MemoSys (Pabinger et al., 2011), to facilitate further improvement and utilization of the reconstruction.

### Flux balance analysis

Once a faithful reconstruction of the metabolic network is available, the corresponding model is ready to be evaluated using constraint-based analysis and other computational methods. A method of choice is FBA, a computational approach that has emerged as a numerically feasible and highly predictive approach to study the properties of large-scale metabolic networks. Crucial to its success, FBA is based on only a few reasonable principles and assumptions, most of which are likely to hold for prokaryotic metabolic networks under many physiological conditions.

To apply FBA, metabolites are subdivided into a set of compounds whose concentrations are assumed to be stationary (internal metabolites) and a set of compounds that may vary over time or are external to the network (external metabolites). Examples of the latter include external nutrients or certain cofactors whose concentrations are assumed to be constant. FBA then makes use of the fact that the set of stationary internal metabolites puts constraints on the stationary flux distributions within the network. In particular, any unchanged metabolite concentration over a given period of time implies that the sum of fluxes producing this metabolite equals the sum of fluxes consuming this metabolite. On the network level, this mass balance constraint reduces the admissible flux space, such that only certain combinations of flux values are feasible. It is noted that the mass balance constraint does not necessarily imply that the system is at steady state. An identical reasoning holds, for example, for oscillating networks—provided that the metabolite concentration is unchanged after a defined period of time, such as a full diurnal cycle. In this case, the flux values must be understood as an integrated flux over the respective period.

Usually, the mass balance constraint itself does not give rise to a specific flux distribution. Within FBA it is therefore assumed that cells have evolved such that their metabolic flux distribution satisfies one or more optimality criteria. For single-celled organisms, such as cyanobacteria, the most common optimality criterion is maximal biomass yield. That is, the metabolic fluxes are assumed to be distributed such that they allow a maximal synthesis yield of all metabolic precursors required for cellular growth. The respective stoichiometric ratios of precursors are provided by the BOF as part of the reconstruction process. It is noted that the optimization process does not consider growth rate, even though the results are usually stated in units of inverse time. In particular, an alternative flux solution that is energetically less favourable but would allow for faster interconversion rates is not selected by FBA (Schuster et al., 2008). An illustration of FBA as a two-step process of constraint and optimization is provided in Fig. 5. A simple example is given in Fig. 6.

Besides the BOF, other cellular objectives can also be explored. In particular, a straightforward optimization of the BOF does not necessarily give rise to a unique flux solution. In this case, additional secondary objectives may be included, such as the minimization of total flux—usually understood as a proxy for the minimization of enzyme synthesis costs required to sustain a particular solution (Holzhütter, 2004). Since its initial formulation, a plethora of refinements and additions have been proposed for FBA. In particular, great strides have been made to include a variety of physicochemical constraints, such as thermodynamic feasibility.
Fig. 5. The principles of flux balance analysis (FBA) as a two-stage constraint optimization: initially, all flux values are constraint by individual upper and lower bounds. Then, assuming stationary conditions, the mass balance constraint is applied, resulting in relationships between intracellular fluxes: the feasible flux cone. A particular solution can be identified by applying optimization criteria, such as the optimization of biomass yield (BOF). The optimal solution is not necessarily unique.

Fig. 6. The principles of FBA: a simple example. The starting point is a reaction network (upper left panel) encoded as a stoichiometric matrix (upper right panel). Given a certain maximal utilization that satisfies the mass balance constraint and results in an optimal yield of biomass formation, as predicted by the biomass objective function, BOF (lower left panel), the problem can be cast into a linear optimization problem and is solved using standard methods of linear programming. A solution for the simple example is provided in the lower right panel. In general, the solution is not unique.
Properties of phototrophic metabolism

While large-scale models of phototrophic and cyanobacterial metabolism are still scarce compared with those of other organisms, a number of theoretical studies of cyanobacterial metabolism have become available over the past years. Preceded by several experimental studies on cyanobacterial metabolism (Yang et al., 2002; Cogne et al., 2003), the first computational analysis based on FBA was performed by Shastri and Morgan (2005), who studied a stoichiometric model of *Synechocystis* sp. PCC 6803 under various growth conditions. The model was later extended by Hong and Lee (2007) and augmented by gene–reaction associations. Subsequently, Fu (2009) presented a genome-scale reconstruction of *Synechocystis* sp. PCC 6803, involving, however, only minimal manual curation. A manually curated model focused on the central metabolism and phototrophic growth was presented by Knoop et al. (2010). Later, two additional models were published (Montagud et al., 2010; Yoshikawa et al., 2011), the latter also conducting a comparison of their own model with the results of *Montagud et al.* (2010) and Knoop et al. (2010).

Interestingly, all models differ with respect to several key pathways. An example is the alleged glyoxylate shunt that is present in all previous reconstructions, except in the reconstruction of Knoop et al. (2010). The presence of such a shunt currently awaits experimental verification. Recent non-stationary flux analysis does not support the presence of the shunt (Young et al., 2011) and there is, to the authors’ knowledge, no convincing experimental evidence of its existence in cyanobacteria. However, the recent discovery of missing enzymes in the citric acid cycle (Zhang and Bryant, 2011) exemplifies that even core metabolic pathways may not be completely described yet. It is noted that none of the models, with the exception of that of Knoop et al. (2010), fulfils current standards of model annotation as requested by MIRIAM (Minimum information requested in the annotation of biochemical models; Le Novère et al., 2005). The lack of such unifying nomenclature makes systematic comparisons unnecessarily difficult.

Similar to genome annotation, large-scale reconstructions of cellular metabolism require continuous revision, extension, and updating. The model originally presented by Knoop et al. (2010) currently includes 608 genes, corresponding to 633 metabolic reactions and 451 metabolites. The focus of the model is a high quality description of cyanobacterial central metabolism, rather than to account for a maximal number of, mostly unverified, reactions. External parameters include maximal photon influx, maximal carbon uptake, as well as possible constraints on the supply of sulphur, nitrate, and inorganic phosphate. Based on a given set of external parameters, the maximal growth yield as well as a, not necessarily unique, flux distribution can be estimated.

Figure 7 shows an estimated flux map for phototrophic growth using a maximal light input of 7.7 mmol photons g DW⁻¹ h⁻¹. The growth yield was maximized with respect to light input using the COBRA toolbox (Schellenberger et al., 2011); other nutrients were not limiting. In addition, cellular maintenance was accounted for by a basal ATPase activity of 0.26 mmol g DW⁻¹ h⁻¹. The flux solution shown in Fig. 7 corresponds to an average doubling time of ~12 h in constant light. Figure 7 also provides an estimated solution for storage-dependent night metabolism. However, the choice of an appropriate objective function under this condition is unclear, as *Synechocystis* sp. PCC 6803 does not exhibit an extensive increase of biomass overnight. This reflects an intrinsic problem of FBA, especially relevant for temporally dynamic systems. Cyanobacterial metabolism is subject not only to external (diurnal) light variation but also to extensive transcriptional remodelling of metabolism via a circadian clock. It is an open and exciting question for modelling of cyanobacteria, how such temporal dynamics can be integrated. In Fig. 7, as a proxy for night metabolism, it is therefore assumed that the metabolism is dominated by glycogen utilization to match cellular maintenance requirements, again implemented as a basal ATP activity. In addition, a small turnover of cellular components is assumed that is proportional to the BOF. The solutions obtained by FBA can be contrasted with measurements of phototrophic flux distributions (Young et al., 2011).

Photorespiration revisited

Importantly, FBA can result in experimentally testable predictions about the functional role of certain reactions. For example, the model of Knoop et al. (2010), and likewise the flux distribution shown in Fig. 7, predicts a non-zero rate for the Rubisco oxygenase (EC 4.1.1.39), a seemingly wasteful side reaction of the Calvin–Benson cycle. This non-zero activity is due to the fact that two enzymes necessary to synthesize the amino acid serine from 3-phosphoglycerate via 3-phosphohydroxypyruvate (EC 2.6.1.52 and EC 3.1.3.3, respectively) are not annotated in the genome of *Synechocystis* sp. PCC 6803. Lacking the conventional synthesis steps, photorespiration provides an alternative pathway to synthesize the amino acids glycine and serine. In particular, using the reconstruction and conditions as defined above, photorespiration leads to a higher yield of serine and glycine per photon than all annotated alternative pathways that could likewise compensate for the absence of the phosphoserine pathway (Knoop et al., 2010).

These findings illustrate several facts about FBA and its utility in the analysis of cellular metabolism: The reconstruction is based on current gene annotation and therefore does not include the phosphoserine pathway. Nonetheless, the respective enzymes may be present in the organism, albeit, as yet, not recognized. Consequences of possible alternative pathways can be tested by adding the respective reactions to the reconstruction – resulting in verifiable predictions about the possible functional role of putative alternative pathways.
Extensions of FBA

While the model of Knoop et al. (2010) and other models have already offered varying degrees of novel insight into the organization of phototrophic metabolism, all studies were as yet limited to fairly standard applications of FBA. However, it must be conceded that the metabolic lifestyle of cyanobacteria grossly differs from the lifestyle of most of those prokaryotic organisms that made applications of FBA popular. Unlike most laboratory or industrial heterotrophic bacteria, cyanobacteria follow a diurnal rhythm that involves drastic changes and re-organizations within its metabolic network. While differences between phototrophic and heterotrophic growth have been frequently studied, the transition from storage-based night metabolism to phototrophic day metabolism exhibits far greater subtlety than...
can currently be captured by conventional FBA. One can envisage a complex temporal coordination of metabolism, such that the biomass function itself is time dependent. Indeed, studies of the cyanobacterial transcriptome (Stöckel et al., 2008) and proteome (Stöckel et al., 2011) suggest a highly coordinated transcriptional programme that is not compatible with current representation of cellular growth by a static BOF. Following the work of Asato (2003, 2006), distinct sequential macromolecular synthesis periods can be defined that are aligned with the diurnal cycle, as well as the cell division cycle (Yang et al., 2010).

**Outlook: towards integrative models**

Cellular metabolism does not operate isolated but is embedded within a network of highly interconnected cellular processes that influence, and are influenced by, the metabolic state of a cell. One of the great challenges of computational modelling is therefore to integrate these diverse processes into a coherent computational framework and to describe cellular interactions on the level of the organism—towards a cyanobacterium in silico.

Processes that directly interact with, or mutually influence, cyanobacterial metabolism include carbon-concentrating mechanisms, the photosynthetic light reactions, the circadian clock, cellular signal transduction, phototaxis, transcriptional regulation, and chromosome organization, as well as the properties of the environment. For several of these processes, computational models already exist, albeit incorporating different levels of detail and representing different levels of reliability. Furthermore, as outlined in the first section, specific processes often require a specific mathematical representation, making the integration of these diverse computational techniques also a considerable technical challenge.

Of primary importance to models of cyanobacterial metabolism is the integration with the photosynthetic light reactions. Photosynthesis itself is reasonably well understood, and several kinetic models have been proposed that describe various aspects of the photosynthetic apparatus in considerable detail. These models are typically based on biophysical principles and aim to incorporate the supramolecular organization and internal states of protein complexes. Detailed models of photosynthesis, such as the model of Jablonsky and Lazar (2008), typically involve a large number of state variables, making integration of such models with metabolism a demanding task. Due to this difficulty, most available kinetic models of the Calvin cycle only employ highly simplified light reactions, such as the direct regeneration of ATP and NADPH by light (Poolman et al., 2000). Nonetheless, efforts to couple both processes already exist (Laisk et al., 2006; Safránek et al., 2011). It should be noted that one of the conclusions drawn from the model of Laisk et al. (2008), a detailed kinetic model that includes light reactions, electron–proton transport, enzymatic reactions, and regulatory functions of C₃ photosynthesis, is that the model is nonetheless insufficient to reproduce the dark–light induction of photosynthesis.

Of similar importance for global regulation of metabolism is the cyanobacterial circadian clock. The cyanobacterial clock consists of a post-translational oscillator (PTO), based on interactions of only three proteins (KaiA, KaiB, and KaiC), coupled to a transcriptional/translational feedback loop (TTFL). As a unique property of the clock, the PTO can be reconstituted in vitro by mixing purified proteins and ATP (Nakajima et al., 2005), and therefore represents a highly attractive guinea pig for the development of detailed computational models. Indeed, a large number of kinetic models were developed over the past years, mostly focusing on intermolecular dynamics of interaction among Kai proteins, entrainment, robustness, and temperature compensation of the clock. In contrast, details of the interaction between the clock, metabolism, and gene expression are largely unknown. Not surprisingly, the clock can be entrained by intracellular levels of ATP (Rust et al., 2011) and, vice versa, the clock seems to exert global control over cellular processes, including global gene expression and the cell cycle (Yang et al., 2010; Johnson et al., 2011). The clock therefore may represent a hinge that allows integration and coordination of different cellular processes.

What should an integrative cyanobacterial model look like? In general, two different approaches are conceivable. On the one hand, a top-down approach can be employed, starting from a black-box view of cellular growth in a photobioreactor that involves only major exchange fluxes. Such models are commonly employed in ecology and biotechnology, and are often based on extensions of Monod’s classic growth equations (Monod, 1949). Top-down models of cyanobacterial growth are usually highly predictive with respect to environmental interactions and do not require detailed knowledge of intracellular states or dynamics. A lucid review on the hurdles and challenges of top-down modelling of phototrophic microorganisms was recently given by Bernard (2011). On the other hand, a modular bottom-up strategy may be employed (Snoep et al., 2006). Here, individual computational models of cellular processes, such as models for the light reactions and the circadian clock, are combined into a coherent whole. The modular approach offers the advantage that individual expert groups can work on, and improve, detailed models of cellular subprocesses, while simultaneously testing the consequence on phototrophic growth by incorporating the respective model into a cellular context. Only such integration will allow an understanding of the multiple feedbacks that arise from interaction of the various subprocesses and define the physiological behaviour of cyanobacteria in their environment. A first example of such an integrative model was recently proposed by Hellwegener (2010). Therein a conceptual model of Synechococcus elongatus PCC 7942 was developed that includes the photosynthetic light reactions, the post-translational oscillator, a minimal metabolism, and a coarse-grained gene expression machinery. The integrated model was used to
investigate the fitness effect of the circadian clock in cyanobacteria *in silico*, again highlighting that complex physiological traits, such as an enhanced fitness of an organism with a resonating clock, cannot be understood in terms of individual molecular subprocesses alone.

It should be emphasized that top-down and bottom-up approaches are not mutually exclusive but should be considered complementary, and can be merged into intermediate coarse-grained representations of cyanobacterial cells. Such coarse-grained representations would allow the gradual replacement of a Monod-type growth rate function. In this sense, both strategies outlined above have strong potential to contribute towards the long-term goal of *in silico* models of cyanobacterial cells and communities in complex environments. Such a step towards a computational ecosystems biology will be instrumental in advancing our understanding of environmental adaptations, complex microbial communities, global carbon and nitrogen cycles, as well as having the potential to facilitate and guide biotechnological applications. Clearly, model integration goes beyond the capabilities of any single research group and will require increased collaboration and exchange between the different areas of research that together define our knowledge of cyanobacterial physiology.

**Acknowledgements**

The authors wish to thank the various colleagues and researcher who provided crucial input on the life of cyanobacteria and aspects of computational modelling, in particular Wolfgang Lockau, Yvonne Zilliges, Ilka M. Axmann, Christian Beck, Thomas Boerner, Wolfgang Hess, Annegret Wilde, Sabrina Hoffmann, Hans Westerhoff, Patrik Jones, and Ladislav Nedbal. Some of the ideas towards cyanobacteria in silico were first discussed at a workshop in Mikulov, Czech Republic, November 10–13, 2010. RS and HK acknowledge financial support by the programm FORSYS-Partner of the German Bundesministerium für Bildung und Forschung (Foerderkennzeichen grant no. 0315274B). RM is grateful for financial support by the SFB 618 Theoretical Biology and by HARVEST Marie Curie Actions Network project 238017.

**References**


Lewis NE, Hixson KK, Conrad TM, et al. 2010. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology 6, 390.


